THE BELL SYSTEM TECHNICAL JOURNAL
Vol. 59, No. 3, March 1980
Printed in U.S.A.

B.S.T.J. BRIEF

Adaptive Aperture Coding for Speech Waveforms—II

By N. S. JAYANT and S. W. CHRISTENSEN

(Manuscript received November 13, 1979)

I. INTRODUCTION

The quality of speech output in adaptive aperture coding[1] has been improved by two refinements: (*i*) a code selection procedure based on formal error minimization rather than observations of aperture crossing,[1] and (*ii*) a simple adaptive low-pass filtering operation based on adjacent sample correlation values, measured on a short-term basis (typically, once every 20 to 30 ms). We describe these refinements with special reference to a 7-code aperture characteristic designed for an *average* output rate of 1.2 bits/sample, and speech inputs sampled at 8 and 12 kHz. At corresponding bit rates (9.6 to 14.4 kb/s), adaptive aperture coding, in conjunction with a first-order adaptive predictor, constitutes a medium-complexity approach in time-domain coding, with an output speech quality that is *less-than-toll* but nevertheless useful in many applications. A natural application of aperture coding is for speech storage where variability of output bit rate is less objectionable than in transmission.

Adaptive aperture coding is a medium-complexity approach to the digitization of slowly changing waveforms. In a recently described[1] procedure, the idea was to form an aperture centered on the last encoded waveform sample and to avoid further encoding until the waveform crossed that aperture. The features of the system that made it applicable to low bit rate digitization of speech were three. The first feature was an arrangement that precluded the need for explicit encoding of aperture crossing times. The second feature was a syllabic adaptation algorithm for varying aperture width in view of the nonsta-

tionarity of the speech waveform. The third feature was the use of the aperture coder in a differential quantization mode, in conjunction with an adaptive first-order predictor. Reference 1 also addressed the variable-output-rate characteristics of aperture coding and suggested that a typical application of the procedure may be for voice storage where variable-rate characteristics would be less objectionable than in real-time communication.

The purpose of this brief is to describe two refinements that have provided improvements in the quality of the speech output from an aperture coder: (*i*) a code selection procedure based on a formal error-minimization rule, rather than observations of aperture crossing as in Ref. 1, and (*ii*) use of a simple, time-varying, low-pass filter based only on short-term adjacent sample correlation, an item of information that is already available in adaptive first-order prediction.

II. APERTURE CODING BASED ON APERTURE CROSSINGS

Refer to the 7-point aperture-(or quantization-) characteristic of Fig. 1a. For the input waveform X shown in the figure, the output will be P3, signifying that an aperture crossing occurred prior to time 3. At

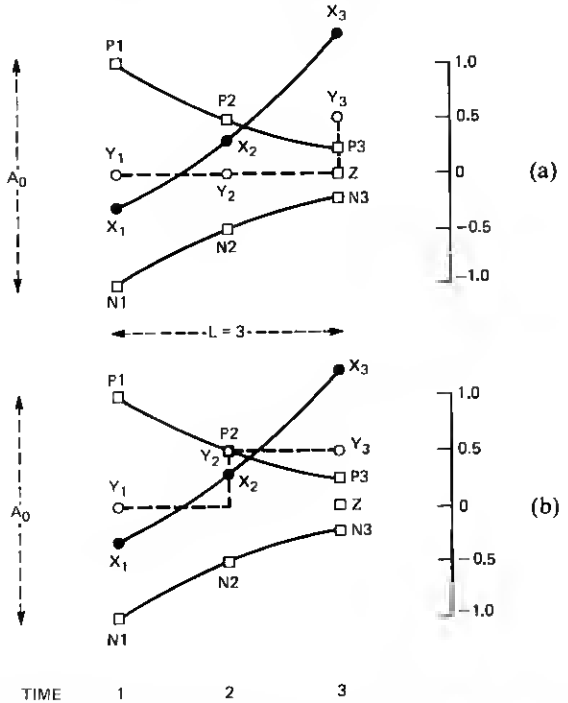

Fig. 1—Adaptive aperture coding based on (a) aperture crossing observations and (b) error-minimizing code selection.

the receiver (or decoder), P3 conveys two items of information: a *timing* information to the effect that the output Y will be updated at time 3, and an *amplitude* information in the sense that the updating magnitude must be positive and *greater than that of P3* on the characteristic. For example, the updating magnitude can be merely that of the code point immediately preceding the transmitted code, and the shape of the aperture characteristic is optimized as in eq. (1) below to make the above amplitude convention appropriate from a quantization-noise viewpoint.[1] The code P3 also implies that the updating is zero at time 2, and the output approximation to the waveform X will therefore follow the dashed line of Fig. 1a. Output codes P1 P2 N1 N2 N3 have corresponding interpretations for waveform updating. For example, if N1 is transmitted, an appropriate negative update occurs at time 1, and the aperture coding procedure repeats with a new aperture characteristic beginning at time 2. We will therefore associate the code N1 (also P1) with a run length of 1. The code P3 as in Fig. 1a implies a run-length of 3 (also equal to the aperture length of 3). The zero code Z occurs if the waveform has not crossed the aperture even at time L. This will be arbitrarily referred to as a run length of $(L + 1)$. When output Code Z is received, the Y-sequence is never updated in the course of a current aperture.

It is desirable for an aperture characteristic to decay exponentially.[1] The width of a single aperture characteristic at phase t is

$$A_0(t) = A_0 \cdot 2^{-J \cdot t}, \tag{1}$$

where the initial aperture width A_0 is adapted by cues derived from a history of the k most recent run-lengths R. Thus for the $(r + 1)$st aperture,

$$\begin{aligned}
A_0^{(r+1)} &= G_1 \cdot A_0^{(r)} \quad \text{if} \quad (\text{ADAPT})^{(r)} = 0 \\
A_0^{(r+1)} &= A_0^{(r)} + G_2 \quad \text{if} \quad (\text{ADAPT})^{(r)} = 1 \\
G_1 &= 1 - \epsilon^2; \quad \epsilon \to 0 \\
(\text{ADAPT})^{(r)} &= 1 \quad \text{if} \quad \sum_{s=1}^{k} R_{r-k} < K \\
&= 0 \text{ otherwise} \\
\text{Max}[A_0^{(r+1)}] &= A_0^{\text{MAX}}.
\end{aligned} \tag{2}$$

Typically, for a 7-code characteristic ($L = 3$) with a 1.2-bit/sample average output rate, $k = 3$ and $K = 7$,[1] and appropriate values for J, G_1, G_2 and A_0^{MAX} are those summarized in Table II. With the above values of k, K', G_1 and G_2, a predominance of high run-length code words (for example, Z) will imply that $A_0^{(r+1)} < A_0^{(r)}$, while a predomi-

nance of low run-length code words (for example, P1 or N1) will imply that $A_0^{(r+1)} > A_0^{(r)}$.

Table I provides a numerical description of the example in Fig. 1a, including data on input X, output Y, and quantization error

$$Q = Y - X. \tag{3}$$

III. APERTURE CODING BASED ON ERROR-MINIMIZING CODE SELECTION

As shown in Fig. 1a, the aperture crossing method of Ref. 1 and Section II suffers from a slope-overloading problem in that rapidly changing inputs are hard to follow. The severity of this problem is a direct result of insisting that the aperture (or quantizer) characteristic should handle both time and amplitude information simultaneously, in an integrated manner. The slope-overload problem is significantly mitigated by a code selection procedure which is based not on aperture crossing per se, but on a minimization of locally averaged variance (or averaged magnitude) of quantization error Q.

In the new procedure, an output code will have a slightly different interpretation. Thus, code P3 will imply updating at time 3, plus an updating amplitude *equal to* that of point P3 on the characteristic. (This is unlike that in Section II where the code P3 implied a crossing prior to time 3, and an updating at time 3 that was *greater* than the value of P3.)

The code selection is now realized by computing an average quantizing error power (or magnitude) for each of the codes in the characteristic, reconstructing tentative Y-waveforms corresponding to each

Table I—Numerical comparison of the two aperture coding methods in Fig. 1

Time n	1	2	3
X_n	−0.3	+0.3	+1.25
	Aperture-Crossing Methods (Output Code = P3)		
Y_n	0.0	0.0	0.5
Q_n	0.3	−0.3	−0.75
Q_n^2	0.09	0.09	0.5625
$E[Q_n^2]$	0.09	0.09	0.2475
	Error-Minimizing Code-Selection Examples Code P2 (Error-Minimizing Code)		
Y_n	0.0	0.5	—
Q_n	0.3	0.2	—
Q_n^2	0.09	0.04	—
$E[Q_n^2]$	0.09	0.065	—
	Code P3		
Y_n	0.0	0.0	0.25
Q_n	0.3	−0.3	−1.0
Q_n^2	0.09	0.09	1.0
$E[Q_n^2]$	0.09	0.09	0.39

of these codes and selecting that waveform and code that corresponds to the least-average quantizing error power (or magnitude). For a 7-code characteristic, the averaging is clearly over 3 (X, Y) pairs for codes Z, P3, and N3, over 2 (X, Y) pairs for P2 and N2 and over one (X, Y) pair, viz., (X_1, Y_1), or codes P1 and N1. This code selection procedure is reminiscent of delayed or tree encoding.[2,3]

For the X-waveform example of Fig. 1, it turns out that the average-error-power-minimizing code is P2, and this leads to the Y-reconstruction shown by the dashed lines of Fig. 1b. Note that this waveform tracks X with much less slope overload than the Y-waveform in Fig. 1a.

Table I compares the effects of choosing codes P2 and P3 numerically using an average error-power criterion $E[Q_n^2]$. The choice of P2 is suggested by its lower final average power error variance (0.065 at time 2), as against the final average for code P3 (0.39 at time 3).

The code selection procedure is formally defined by

$$\text{Select code } C_m \quad \text{if} \quad E[Q^2]\,|\,_{C_m} < E[Q^2]\,|\,_{C_p} \tag{4}$$

for all codes C_p, $p \neq m$, with

$$E[Q^2]\,|\,_{C_p} = \left[\frac{1}{l(C_p)} \sum_{u=1}^{l(C_p)} (X_u - Y_u)^2\right],$$

where $l(C_p)$ is the number of samples up to and including the point where C_p appears on the characteristic. Clearly,

$$\text{Max}[l(C_p)] = L.$$

The superiority of the error-minimizing approach has been confirmed by extensive computer simulations which involved different input samples and two sampling frequencies, 8 and 12 kHz. Design parameters were separately optimized, as shown in Table II. Signal-to-noise comparisons are provided in Table III, where the S/N ratios are signal-to-quantization error variance ratios. SNR is the conventional long-time averaged ratio expressed in decibels, while the segmental[4]

Table II—Desirable designs for two aperture coding systems with $L = 3$ and average output bit rate of 1.2 bits/sample. Values of G_2 and A_0^{MAX} are appropriate for a maximum speech amplitude of ±32000

Sampling Frequency	8 kHz				12 kHz			
Coding based on	J	G_1	G_2	A_0^{MAX}	J	G_1	G_2	A_0^{MAX}
Aperture Crossing	0.5	0.99	50	8000	0.5	0.99	30	5000
Error Minimizing Code Selection	1.0	0.95	250	8000	1.0	0.95	80	5000

Table III—Objective performance comparisons for the two aperture coding methods used in conjunction with first-order adaptive prediction. *F*-subscripts refer to ratios after adaptive filtering. All S/N ratios are in decibels

Sampling Frequency	8 kHz				12 kHz			
Coding based on	SNR	SNRSEG	SNR_F	$SNRSEG_F$	SNR	SNRSEG	SNR_F	$SNRSEG_F$
Aperture Crossing	7.3	8.9	7.6	9.3	9.3	11.7	10.5	12.9
Error Minimizing Code Selection	8.8	10.9	8.7	11.6	11.4	14.4	12.0	15.9

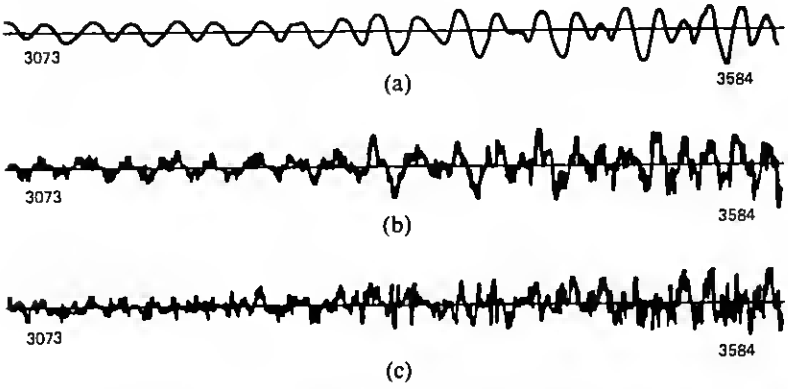

Fig. 2—Waveforms of (a) input speech and quantizing error in adaptive sperture coding based on (b) aperture crossing and (c) error-minimizing code selection [error waveforms (b) and (c) have been msgnified by a factor of 5].

ratio SNRSEG is the average value of short-term (average over, say, 16 ms) S/N values each of which is expressed in decibels prior to the averaging—a procedure that better reflects the rendition of low-level waveform segments.

Finally, Fig. 2 illustrates how the error waveform also tends to be more noise-like (less speech-correlated) if error-minimizing code selection is employed, signifying a perceptual improvement in the aperture-coding process.

IV. ADAPTIVE LOW-PASS FILTERING OF OUTPUT SPEECH

At bit rates in the range of 9.6 to 16 kb/s use, we have found that it is very desirable to smooth the output of an aperture coder by some kind of an adaptive low-pass filter. In fact, even a sloppy low-pass filter characteristic such as

$$(YF)_r = B \cdot (YF)_{r-1} + (1 - B) \cdot Y_r, \tag{4}$$

where YF represents a filtered version of Y, is quite effective provided B is appropriately adaptive. We have studied the sloppy procedure (4) at some length[5] because it involves a single parameter B which can be meaningfully related to the value of the local adjacent sample correlation C in the speech waveform, a parameter that is already available in first-order adaptive prediction (with time-varying coefficient $h_1 = c$).

An interesting adaptation approach is that exemplified by

$$B(C) = P \cdot C + Q \tag{5}$$

with typical (P, Q) settings of (0.4, 0.4) or (0.3, 0.3). Note that the basic idea is to provide the most smoothing (greatest B) for the very slowly varying ($C \rightarrow 1$) waveform segments of voiced speech.

When $C \rightarrow 1$, the local bandwidth tends to be low (much less than half the sampling rate) and low-pass filtering of the output is clearly very effective for quantizing noise rejection.

The gains due to adaptive filtering as described in (4) are illustrated by the objective improvements, shown by subscripts F, in Table III, while design principles for (4) and (5) are discussed elsewhere.[5] Noise reduction with the first-order filter approach entails in general a concomitant loss of speech crispness, and this can be avoided if one is willing to employ sharper adaptive filters, a procedure that will also be discussed separately[6] in the context of a delta modulation coder.

REFERENCES

1. N. S. Jayant and S. W. Christensen, "Adaptive Aperture Coding of Speech Waveforms—I," B.S.T.J., *58*, No. 7 (Sept. 1979), pp. 1631–1645.
2. C. C. Cutler, "Delayed Encoding: Stabilizer for Adaptive Coders," IEEE Trans. Commun. Technol. *COM-19*, No. 6 (Dec. 1971), pp. 898–904.
3. N. S. Jayant and S. W. Christensen, "Tree Encoding of Speech Using the (M, L)-Algorithm and Adaptive Quantization," IEEE Trans. Commun., *COM-26* (Sept. 1978), pp 1376–1379.
4. P. Noll, "Adaptive Quantization in Speech Coding Systems," Int. Zurich Seminar on Digital Communications, March 1974, pp B3.1-B3.6.
5. N. S. Jayant, "Adaptive Low-Pass Filtering Based on Short Term Values of Adjacent Sample Correlation," unpublished work.
6. J. Smith, J. B. Allen, and N. S. Jayant, "Adaptive Delta Modulator Gains due to Time-Varying Filtering and Sampling," unpublished work.